



UNIVERSITAT POLITÈCNICA  
DE CATALUNYA  
BARCELONATECH



Capítulo 4:

# Probabilidad, riesgo, *odds* y tasa

Erik Cobo, José A. González y Pilar Muñoz

Jordi Cortés, Rosario Peláez, Marta Vilaró, Hector Rufino y Nerea Bielsa

Septiembre 2014



## Probabilidad, riesgo, *odds* y tasa

<b>Presentación .....</b>	<b>2</b>
<b>1. Probabilidad .....</b>	<b>3</b>
1.1. Introducción a la probabilidad .....	3
1.2. Probabilidad condicionada .....	4
1.3. Independencia .....	5
1.4. Aplicación de la probabilidad al diagnóstico .....	6
1.5. Teorema de Bayes * .....	10
<b>2. Riesgos y tasas.....</b>	<b>12</b>
2.1. Riesgo.....	12
2.2. <i>Odds</i> o momio .....	13
2.3. <i>Odds</i> y Bayes .....	16
2.4. Tasa .....	17
2.5. Incidencia y prevalencia.....	19
<b>3. Asociación .....</b>	<b>20</b>
3.1. Diferencia de riesgos.....	20
3.2. Número necesario de casos tratados .....	21
3.3. Riesgo relativo .....	22
3.4. <i>Odds</i> ratio o razón de momios .....	23
3.5. Razón de tasas ( <i>Hazard ratio</i> o HR) .....	24
3.6. Relación entre los 3 cocientes de riesgos: RR, HR y OR * .....	25
3.7. Relación entre HR y las medianas de los tiempos hasta el evento. ....	26
<b>Soluciones a los ejercicios .....</b>	<b>27</b>

\* Indica tema más avanzado que conviene mirar pero no es crucial para los ejercicios.

## Presentación

No es ningún secreto que muchos fenómenos son inciertos. A las mentes más deterministas les gusta pensar que esta incertidumbre es el resultado de la falta de conocimiento: si se supieran cuáles son todas las fuerzas que actúan en un momento dado y cómo lo hacen, se podría predecir exactamente el resultado.

El modelo estadístico divide a estas causas en dos grandes grupos: el primero lo forman un número limitado con suficiente entidad para que su efecto pueda ser estimado y modelado de forma clásica. En cambio, el segundo grupo es ilimitado y con influencia reducida: son tantas, y con efectos tan pequeños, que la única manera de modelar el resultado de su influencia es mediante las teorías de combinatoria y de probabilidad.

En este capítulo se introducen, con la ayuda de ejemplos, conceptos necesarios para interpretar los resultados de una prueba diagnóstica, la capacidad predictiva de un indicador, el concepto de riesgo o las medidas de comparación de riesgos. Para aquellas situaciones en las que varía el tiempo de seguimiento, se introducen las tasas.

**Contribuciones:** La versión original de marzo 2013 descansa en el libro de Bioestadística para No estadísticos de Elsevier de EC, JAG y PM, editada por JC y revisada por MV y RP y HR; y la de septiembre de 2104 por NB y EC.

# 1. Probabilidad

## 1.1. Introducción a la probabilidad

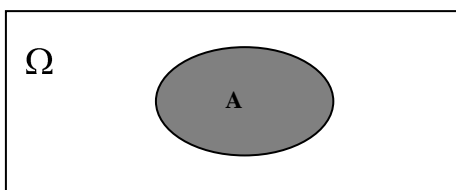


### Definición

La **probabilidad**, desde el punto de vista subjetivo es la expectativa de que ocurra un suceso.

**Lectura:** Varias definiciones de probabilidad han sido propuestas. La primera definía la probabilidad como el “cociente entre casos favorables y casos posibles”. Dada su circularidad, (exige equiprobabilidad) enseguida se propuso sustituirla por “la frecuencia de aparición de un suceso”. En el fondo, ambas definiciones son métodos para asignar valores a la probabilidad, es decir para conocer sus valores, pero no definiciones formales de lo que es en sí misma la probabilidad. En el siglo pasado se propusieron, casi al mismo tiempo, dos definiciones muy interesantes. Una de ellas, abstracta y [axiomática](#), permitió un desarrollo formidable de toda la teoría de probabilidad. La otra, [subjetiva](#), expresa en términos de probabilidad el grado de creencia en una afirmación científica, y también la expectativa de (la ocurrencia de) un suceso. Su [influencia](#) en los conceptos de ciencia y de conocimiento está por desarrollarse plenamente.

La **Figura 1.1** representa la probabilidad del suceso  $A$ ,  $P(A)$ , dónde su valor concreto es directamente el cociente entre el área sombreada ( $A$ ) y el área total ( $\Omega$ ).



**Figura 1.1** Probabilidad del suceso  $A$

**Ejemplo 1.1:** en 1978 se registraron 66 muertes por leucemia en una población de 890575 individuos. Si se estima la probabilidad de morir por leucemia mediante la proporción (estimador de la probabilidad), se obtiene un riesgo de 74 muertes por millón de habitantes.

$$P = \text{muertes por leucemia} / \text{población total} = 66 / 890575 = 0.0000741$$



### Definición

La **probabilidad de un evento** es el número de casos favorables dividido por el número de casos posibles.

Como que las unidades del numerador son las mismas que las del denominador, éstas se cancelan y la probabilidad no tiene unidades de medida.



### Recuerde

Si el denominador de la medida de frecuencia es el número total, hablamos de probabilidad.

Si representamos por “no A” al complementario de A:  $P(\text{no } A) = 1 - P(A)$ .

## 1.2. Probabilidad condicionada

Una gran cantidad de factores pueden modificar esta probabilidad. Para considerarlas es preciso realizar la siguiente definición.



### Definición

**Probabilidad del suceso A condicionado al suceso B** es la probabilidad de aparición del suceso A sabiendo que B es cierto.

Se representa por  $P(A|B)$ .

**Ejemplo 1.2:** La probabilidad de nacer con los ojos azules en España (suceso A) es de aproximadamente  $1/6$ , pero si se conoce que ambos progenitores tienen los ojos azules (suceso B), está probabilidad condicionada se incrementa hasta 0.99 (99%).

La probabilidad condicionada se obtiene a partir del cociente entre la probabilidad de observar a la vez ambos sucesos ( $A \cap B$ ) y la del suceso B.



### Definición

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\text{Probabilidad de ser, a la vez, A y B}}{\text{Probabilidad de ser B}}$$

La probabilidad de A condicionada a B,  $P(A|B)$ , estudia sólo los casos que cumplen la característica B.

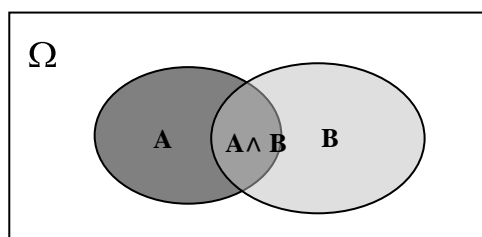


Figura 1.2 Probabilidad del suceso A condicionado al suceso B



### Recuerde

$P(A|B)$  representa la probabilidad del suceso A ‘dentro’ del total de casos que cumplen B.

$P(A \cap B)$  representa la probabilidad de que ocurran simultáneamente los sucesos A y B.



### Ejercicio 1.1

¿Qué sería en el gráfico  $P(B|A)$ ? ¿Cuál sería su definición?

La probabilidad condicional no posee la propiedad conmutativa, es decir, no se cumple que  $P(A|B) = P(B|A)$ .

**Ejemplo 1.3:** [Stephen Senn](#) recuerda la probabilidad de que un católico sea Papa es muy baja, pero la de que el Papa sea católico es muy alta.

**Ejemplo 1.4:** la probabilidad de que alguien que ha parido sea mujer es altísima, quizás 1, pero la probabilidad de que una mujer haya parido no es tan alta.

**Ejemplo 1.5:** Podemos interpretar que la probabilidad de contraer cáncer (C) de pulmón si se es fumador (F) es elevada  $P(C|F)$ ; pero la probabilidad de haber sido fumador si se tiene cáncer  $P(F|C)$  es mayor.

**Nota técnica:** como hay más gente que fuma que con cáncer,  $P(F) > P(C)$ , y como ambos tienen el mismo numerador,  $P(F \cap C)$ ,  $P(F|C) > P(C|F)$ .



### Ejercicio 1.2

Sea C tener cáncer de Mama y M ser mujer. ¿Qué es mayor,  $P(M|C)$  o  $P(C|M)$ ?

## 1.3. Independencia

La idea de eventos independientes está ligada a la de la información que uno aporta sobre el otro: A y B son independientes cuando la probabilidad de A es la misma indiferentemente de lo que pase con B (B no informa sobre la probabilidad de A).

$$\text{Si A y B son independientes, entonces } P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A).$$

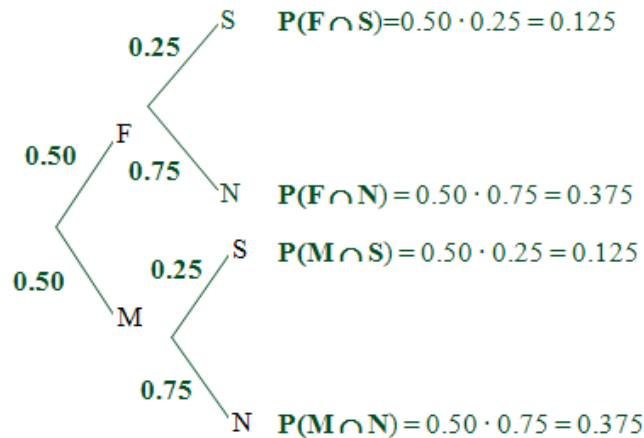
Y viceversa, que pase A no cambia la expectativa de B:

$$\text{Si A y B son independientes, entonces } P(B|A) = \frac{P(A \cap B)}{P(A)} = P(B).$$

En resumen independencia aplicado a dos eventos tiene doble implicación:

- si  $P(A \cap B) = P(A) \cdot P(B)$ , entonces A y B son independientes
- si A y B son independientes, entonces  $P(A \cap B) = P(A) \cdot P(B)$

**Ejemplo 1.6:** Miopía (Sí/No) y Género (Masculino/Femenino) son independientes:



### Ejercicio 1.3

Explique porqué son independientes: ¿qué números compara para decirlo?

## 1.4. Aplicación de la probabilidad al diagnóstico

Por simplicidad, suponga que debe diagnosticar una sola enfermedad con dos únicos posibles estados, enfermo (E) y sano (S); y que se dispone de un único indicador con dos posibles valores, positivo (+) y negativo (-).

**Lectura:** [Bossuyt, Reitsma, Bruns](#) ponen como ejemplo de una buena definición de objetivos en el resumen de un original: “*Purpose: To determine the sensitivity and specificity of computed tomographic colonography for colorectal polyp and cancer detection by using colonoscopy as the reference standard.*”



### Recuerde

La primera dificultad es definir las variables: (1) la referencia o *gold standard* mide “perfectamente y sin error”, enfermo/sano; y (2) el resultado del test, indicador o prueba que se desea estudiar.



### Definición

**Sensibilidad (Sens):** tendencia o propensión de los enfermos a dar positivo (en esta prueba).

**Especificidad (Esp):** tendencia o propensión de los sanos a dar negativo.

**Valor predictivo positivo (VP+):** confianza o credibilidad de un resultado positivo (de esta prueba).

**Valor predictivo negativo (VP-):** confianza o credibilidad de un resultado negativo.

**Ejemplo 1.7:** Una prueba sería muy-muy sensible si, aplicada a un conjunto de enfermos, casi el 100% dan positivo. Y tendrá un elevado valor predictivo positivo si casi el 100% de los que dan positivo están realmente enfermos. ¡Parecen lo mismo, pero no lo son!

La probabilidad condicionada permite definir estos conceptos formalmente, aclarando diferencias y similitudes. De hecho, aparecen cuatro probabilidades condicionadas:  $P(E|+)$ ,  $P(+|E)$ ,  $P(-|S)$  y  $P(S|-)$ .



#### Ejercicio 1.4

Relaciona estas 4 probabilidades condicionadas con los 4 conceptos anteriores de sensibilidad, especificidad y valores predictivos positivo y negativo.

En la sensibilidad y en la especificidad, el condicionante, o punto de salida, es la realidad (enfermo o sano); mientras que el condicionado, o punto de llegada, es el indicador (positivo o negativo). Así, ambas van de la realidad al resultado: siguen el planteamiento *racionalista* de ir de la causa a la consecuencia. Los valores predictivos, en cambio, van al revés: condicionan o parten del indicador y se preguntan por el estado real del paciente. Por ello contestan la pregunta *empirista* del diagnóstico: dado que este paciente ha presentado estos síntomas, signos e indicadores, ¿está enfermo? O mejor: ¿qué probabilidades tiene de estar realmente enfermo sabiendo que ha dado positivo,  $P(E|+)$ ? ¿O de estar sano si ha dado negativo,  $P(S|-)$ ?

**Ejemplo 1.8:** la sensibilidad, la especificidad y los valores predictivos de la prueba diagnóstica cuyos resultados resume la Tabla 1.1, son:

	+	-	Total	
Enfermo	94	38	132	$Sens = P(+ E) = 94/132 \approx 0.712 = 71.2\%$
Sano	215	653	868	$Esp = P(- S) = 653/868 \approx 0.752 = 75.2\%$
Total	309	691	1000	$VP+ = P(E +) = 94/309 \approx 0.304 = 30.4\%$
				$VP- = P(S -) = 653/691 \approx 0.945 = 94.5\%$

**Tabla 1.1.** Probabilidades diagnósticas en una muestra con un 13.2% de enfermos. Sensibilidad y Especificidad son los porcentajes de fila, pero los VP son los porcentajes de columna.

A pesar de que los valores de sensibilidad y especificidad son muy similares, los valores predictivos se alejan considerablemente entre sí: la probabilidad de que un paciente que dé negativo esté sano (VP-) es alta, pero la probabilidad de que un paciente que dé positivo esté



enfermo (VP+) es baja. La razón es muy simple: hay más sanos (86.8%) que enfermos (13.2%) y al calcular los VP, que son proporciones sobre el total de la columna, la segunda fila ‘pesa’ más que la primera.



### Ejercicio 1.5

Reconstruya la siguiente tabla si sensibilidad = especificidad = 100%. ¿Se cumple que ambos valores predictivos son del 100%?

	+	-	Total
Enfermo			132
Sano			868
Total			1000



### Recuerde

Si la sensibilidad y la especificidad son ambas del 100%, no hay duda: un caso positivo está enfermo y un negativo está sano.



### Ejercicio 1.6

Mirando a la tabla construida en el ejercicio anterior, para que el valor predictivo positivo sea perfecto ( $VP+ = 100\%$ ), ¿qué necesita que sea del 100%, la sensibilidad o la especificidad?



### Recuerde

Sensibilidad perfecta implica VP- perfecto (100%).

Especificidad perfecta implica VP+ perfecto (100%).



### Ejercicio 1.7

Intente explicar con sus propias palabras lo que miden la sensibilidad, la especificidad y ambos valores predictivos.

### Ejercicio 1.8

Un signo es patognomónico si su presencia asegura definitivamente la enfermedad: ¿alguna de las probabilidades diagnósticas es del 100%?

**Ejemplo 1.9:** Tal y como puede observarse en la Tabla 1.2, si cambiamos  $P(E)$  de 0.132 a 0.75, aunque mantengamos la sensibilidad y la especificidad iguales a las anteriores, los valores predictivos cambian.

	+	-	Total	
Enfermo	534	216	750	$Sens = P(+ E) = 534/750 \approx 0.712 = 71.2\%$
Sano	62	188	250	$Esp = P(- S) = 188/250 \approx 0.752 = 75.2\%$
Total	596	404	1000	$VP+ = P(E +) = 534/596 \approx 0.896 = 89.6\%$
				$VP- = P(S -) = 188/404 \approx 0.465 = 46.5\%$

**Tabla 1.2** Probabilidades diagnósticas en un entorno muy especial, con un 75% de enfermos

No olvide cuál es la proporción de enfermos en la población origen de los datos. Si es muy baja, la proporción de enfermos seguirá siendo baja después del resultado de la prueba, lo que llevará a un valor predictivo positivo (proporción de enfermos en los positivos) bajo y a un valor predictivo negativo (proporción de sanos en los negativos) alto.

Un observador poco atento podría olvidar las condiciones del entorno y esperar que, si un 70% de los enfermos dan positivo, aproximadamente un 70% de los positivos estarían enfermos.



### Ejercicio 1.9

Repita los cálculos, suponiendo que dispone de 2 muestras de 1000 casos, una para enfermos y otra para sanos. (Tabla 1.3)

	+	-	Total
Enfermo	712	288	1000
Sano	248	752	1000
Total	960	1040	

**Tabla 1.3** Probabilidades diagnósticas en dos muestras de 1000 casos

En el ejercicio anterior ambas filas pesan lo mismo ya que ambas filas se han fijado ‘artificialmente’, por lo que estos valores predictivos sólo aplicarían a la irreal situación en la que ambas filas fueran igual de frecuentes.



### Recuerde

Muchos diseños se basan en 2 muestras (enfermos y sanos) obtenidas por separado. Como sus números de casos los decide el investigador, no permiten calcular directamente los valores predictivos.

Cuando el número de enfermos y de sanos está fijado por muestreo, el cálculo de los VP debe usar el método explicado en el siguiente apartado.

### 1.5. Teorema de Bayes \*

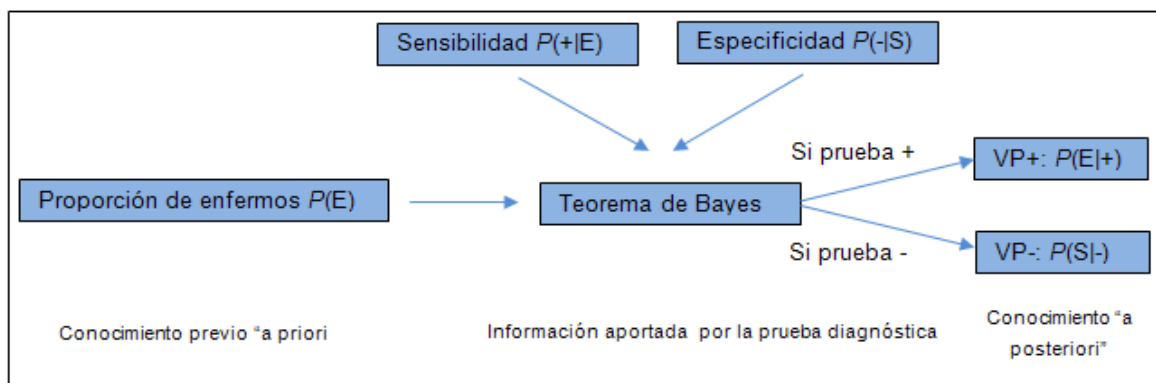
La sensibilidad y la especificidad son el resultado de un mecanismo fisiológico o patológico y, por tanto, suelen ser extrapolables de una población de pacientes a otra. En cambio, los valores predictivos dependen de la frecuencia de la enfermedad en la población, que puede variar mucho de un entorno a otro.



#### Recuerde

La sensibilidad y la especificidad quizás puedan ser transportables de una población a otra, pero los valores predictivos seguro que no, ya que dependen de la frecuencia de la enfermedad en cada población.

**Ejemplo 1.10:** suponga dos poblaciones diferentes: (1) prisioneros con hábitos de riesgo y (2) profesionales sanitarios. Asumamos que, en la primera población, la frecuencia de SIDA es mucho mayor. Y supongamos que 2 casos, uno de cada entorno han dado positivo. ¿Tienen la misma probabilidad de tener el SIDA?



**Figura 1.3** Esquema del Teorema de Bayes

El teorema de Bayes permite resolver este problema, ya que combina ambas piezas de información (la proporción de enfermos de la población con el rendimiento del indicador, resumido en sus valores de sensibilidad y especificidad) para obtener los valores predictivos.



#### Definición.

$$P(B|A) = \frac{P(A|B) P(B)}{P(A)}$$

El teorema de Bayes permite invertir condicionante y condicionado: a partir de  $P(A|B)$  se obtiene  $P(B|A)$  y por tanto, a partir de  $P(+|E)$  se llega a  $P(E|+)$ .

El teorema de Bayes permite invertir condicionante y condicionado: a partir de  $P(A|B)$  se obtiene  $P(B|A)$  y por tanto, a partir de  $P(+|E)$  se llega a  $P(E|+)$ .

**Ejemplo 1.11:** Sea  $A$  = “Estar infectado por el virus de la Gripe A” y  $B$  = “Tener fiebre superior a  $38^{\circ}\text{C}$ ”, entonces dicha fórmula permite conocer la proporción de Gripe A entre los que tienen fiebre, o  $P(A|B)$ , a partir de la proporción de fiebre entre los de Gripe A, o  $P(B|A)$ ; y la prevalencia, o  $P(A)$ .

**Nota técnica:** demostración a partir de la definición de probabilidad condicionada

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(A|B) \cdot P(B)$$

Y entonces 
$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B) \cdot P(B)}{P(A)}$$

**Nota:** la fórmula de Bayes se puede expresar también mediante la descomposición del denominador  $A$  en los dos sucesos  $A \cap B$  y  $A \cap (\text{no } B)$  que lo componen .

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)} = \frac{P(A|B) \cdot P(B)}{P(A \cap B) + P(A \cap \text{no } B)} = \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|\text{no } B) \cdot P(\text{no } B)}$$



### Recuerde

Una expresión alternativa de la fórmula de Bayes es:

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)} = \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|\text{no } B) \cdot P(\text{no } B)}$$

**Ejemplo 1.12:** al aplicar las sensibilidades y especificidades obtenidas en el Ejemplo 1.9 a un entorno con  $P(E) = 0.132$ , el valor predictivo positivo será:

$$P(E|+) = \frac{P(+|E) \cdot P(E)}{P(+|E) \cdot P(E) + P(+|S) \cdot P(S)} = \frac{0.712 \cdot 0.132}{0.712 \cdot 0.132 + 0.248 \cdot 0.868} \approx 0.304$$



### Ejercicio 1.10

Compruebe que sabe calcular el valor predictivo negativo.



### Recuerde

El teorema de Bayes, a partir de la proporción de enfermos, la sensibilidad y la especificidad proporciona los valores predictivos.

El teorema de Bayes permite ir actualizando la información: su fórmula “mezcla” la información previa, disponible “a priori”  $P(E)$ , con los nuevos resultados (+ ó -).



### Ejercicio 1.11

Volvamos al ejemplo del resultado positivo del SIDA en un recluso con hábitos de riesgo y en un colega. Suponga que en el primer colectivo, la proporción de SIDA es un 80% y en el segundo un 1%. Además, tanto sensibilidad como especificidad valen 0.95. Calcule VP+ en ambos casos.



### Recuerde

Información a priori + nueva información = Información a posteriori.

## 2. Riesgos y tasas

### 2.1. Riesgo



### Definición

En sentido amplio, **riesgo** es la probabilidad de que algo desfavorable (un evento negativo) ocurra.

En investigación clínica, indica la probabilidad de que aparezca un fenómeno adverso concreto, quizás tras una actividad, intervención o exposición..

**Ejemplo 2.1:** según [Baños et al \(3\)](#), la [Food and Drug Administration \(FDA\)](#) considera lícito someter a los voluntarios de estudios sin beneficio terapéutico, a un riesgo “mínimo o insignificante”, que define como una probabilidad de entre 1 y 100 por mil de sufrir una complicación menor; o de entre 10 y 1000 por millón de sufrir una grave.

**Nota:** En Medicina, riesgo y probabilidad suelen ser sinónimos. Pero la definición de [función de riesgo](#) en teoría de la decisión incluye también las consecuencias (“pérdidas”). Así, aunque (supongamos) las probabilidades de padecer en cierto tiempo gripe y cáncer sean iguales, para matemáticos, estadísticos y economistas el “riesgo” de la segunda es mayor. En esta línea, la definición anterior de “mínimo o insignificante” por la FDA, baja la frecuencia cuando sube la gravedad. Sin embargo en lo que queda de capítulo, usaremos el término riesgo por su acepción usual en Medicina y Epidemiología, es decir, como sinónimo de frecuencia, sin considerar las consecuencias.

Veamos su cálculo en unos datos sencillos.

**Ejemplo 2.2:** Suponga que una enfermedad (Y) y su Factor de Riesgo (X) sólo pueden tomar dos valores: presente (+) y ausente (-). Así, Y+ representará tener la enfermedad; y X-, que no está expuesto al factor de riesgo.

En la tabla 2.1 puede leerse que de 1000 casos, 15 presentaban la enfermedad, de los que 7 estaban expuestos y 8 no.

	Y+	Y-	Total
X+	7	125	132
X-	8	860	868
Total	15	985	1000

**Tabla 2.1** Presencia de la enfermedad (Y) y del factor de riesgo (X) en 1000 casos

Riesgo:  $P(Y+) = 15 / 1000 = 0.015$

Riesgo en los Expuestos:  $P(Y+|X+) = 7 / 132 \approx 0.053$

Riesgo en los No-Expuestos:  $P(Y+|X-) = 8 / 868 \approx 0.009$



### Ejercicio 2.1

[Carmina R. Fumaz](#) compara 2 grupos de pacientes tratados, uno con Efavirenz (EFV, n=51) y otro con Inhibidores de la Proteasa (PI, n=49), habiendo observado respectivamente, 36 y 9 acontecimientos adversos relacionados con el sistema nervioso central. Construya la tabla 2x2 y calcule los riesgos respectivos.

## 2.2. Odds o momio

Los países de tradición anglosajona usan una forma alternativa para expresar resultados inciertos. Si la probabilidad expresa “casos a favor divididos por todos los casos posibles”, la *odds* habla de “casos a favor divididos por casos en contra”.

**Ejemplo 2.3:** así, mientras nosotros diríamos que cierto caballo tiene 7 números sobre (un total de) 8 de ganar una carrera, los anglosajones suelen decir que los números de este caballo están 7 a favor frente a 1 en contra.



### Recuerde

Si el denominador de la medida de frecuencia es el número de “casos en contra”, hablamos de *odds*.

**Ejemplo 2.4:** en los ambientes de apuestas (pelota vasca, carreras de galgos,...) se dice, por ejemplo, que las apuestas por el pelotari A están 7 momios a 1. Su gran ventaja es que

facilita el cálculo del premio: además de recuperar su inversión, los que apuestan por A obtendrían 7, pero los que lo hagan por B, 1/7 (además de recuperar su inversión).



### Recuerde

Usar odds permite calcular rápidamente el momio o beneficio potencial de una apuesta.

**Lectura:** Wikipedia en “[razón de momios](#)” explica otros intentos de traducir *odds*.



### Definición

La *odds* de A es la probabilidad de que se presente el suceso A dividida por la probabilidad de que no se presente A.

$$Odd A = \frac{P A}{P \text{ no } A}$$



### Ejercicio 2.2

¿Cuánto valen la probabilidad y la *odds* de sacar un “3” en el lanzamiento de un dado? En un juego de apuestas “justo” ¿Cuánto debería pagarse si saliera un 3? (¿A cuánto estará el momio?)

**Ejemplo 2.5:** Veamos un ejemplo de cálculo sencillo.

*Odds* en los Expuestos:  $O(Y+ | X+) = 7 / 125 \approx 1 / 18 \approx 0.056$

*Odds* en los No-Expuestos:  $O(Y+ | X-) = 8 / 860 \approx 1 / 107 \approx 0.0093$

En los expuestos, la enfermedad aparece en 1 caso por cada 18 que no aparece. En cambio, en los no-expuestos, la enfermedad aparece en 1 caso por cada 107 en los que no.

**Nota:** si la probabilidad de enfermedad es muy pequeña, la probabilidad de sano será muy próxima a 1, por lo que la *odds* tendrá un valor muy similar a la probabilidad:

$$Odds(enfermo) = \frac{P(enfermo)}{P(sano)} \approx \frac{P(enfermo)}{1} = P(enfermo)$$



### Recuerde

En el caso de enfermedades “raras”, riesgo y *odds* dan resultados similares.

**Ejemplo 2.5 (continuación):** La *odds* en los expuestos vale 0.056 [ $O(Y+|X+) = 7/125 \approx 0.056$ ], muy similar al 0.053 anterior del riesgo en los expuestos.

En los no expuestos la similitud es aún mayor: la odds vale 0.0093 y el riesgo 0.0092.



### Ejercicio 2.3

El dolor lumbar o la gripe son enfermedades comunes en el sentido de que a lo largo de la vida es fácil padecerlas al menos en una ocasión. Pongamos que sus probabilidades respectivas son 0.5 y 0.8. Calcule sus *odds*. La esclerosis Múltiple, en cambio, es muy poco frecuente. Pongamos que la probabilidad de padecerla a lo largo de la vida vale 0.001 (uno por mil). Calcule la *odds*. Interprete.

De la misma forma que los valores predictivos no podían calcularse si el diseño implicaba dos muestras, una de enfermos y una de sanos, el siguiente ejercicio muestra que según el diseño del muestreo tampoco pueden calcularse los riesgos.



### Ejercicio 2.4

Se ha recogido en 1000 casos la exposición al tabaco (fumador: X+ y no fumador: X-) y su evolución posterior (bronquitis: Y+ y no bronquitis: Y-). Calcule riesgos y odds de la tabla 1.

Repita los cálculos en con la tabla 2, con los datos de dos muestras de 1000 casos, una de fumadores y otra de no fumadores.

Repita una vez más, suponiendo ahora que las dos muestras de 1000 casos corresponden, una a bronquíticos y otra a no bronquíticos (Tabla 3).

Tabla 1	Y+	Y-	Total
X+	94	38	132
X-	215	653	868
Total	309	691	1000

Tabla 2	Y+	Y-	Total
X+	712	288	1000
X-	248	752	1000
Total	960	1040	2000

Tabla 3	Y+	Y-	Total
X+	304	55	359
X-	696	945	1641
Total	1000	1000	2000



### Recuerde

Si se fija por diseño el número total de casos con la enfermedad y el número de casos sin la enfermedad, ya no puede calcular ni los riesgos ni las *odds* de desarrollar la enfermedad (ni en los expuestos, ni en los no expuestos).



### 2.3. Odds y Bayes

Trabajar con *odds* en lugar de con probabilidades simplifica Bayes: la *odds a posteriori* de enfermo a sano es la *odds a priori* por la *razón de verosimilitud* (*likelihood ratio*).

La razón de verosimilitud de un resultado positivo (RV+) es el cociente de las 2 probabilidades de positivo condicionadas (por enfermo y por sano).



#### Recuerde

Odds a posteriori = Odds a priori \* razón de verosimilitud

$$RV+ = P(+|E) / P(+|S)$$

**Nota técnica:** utilizando probabilidades condicionadas se obtuvo que:

$$P(E | +) = P(+ | E)P(E) / P(+) \text{ y } P(S | +) = P(+ | S)P(S) / P(+)$$

Recordando la definición de la *odds*,  $Odds(E) = P(E) / P(S)$  y aplicándola a las probabilidades condicionadas anteriores, se pueden obtener las *odds* “a posteriori” entre enfermo (E) y sano (S), una vez conocido el resultado positivo de la prueba:

$$\underbrace{\frac{P(E | +)}{P(S | +)}}_{\text{Odds a posteriori}} = \frac{P(+ | E)P(E) / P(+)}{P(+ | S)P(S) / P(+)} = \underbrace{\frac{P(+ | E)}{P(+ | S)}}_{\text{Razón verosimilitud (RV)}} \cdot \underbrace{\frac{P(E)}{P(S)}}_{\text{Odds a priori}}$$

**Ejemplo 2.6:** A partir de los datos del Ejemplo 1.8, RV+ (razón de verosimilitud) vale

$$\frac{P(+|E)}{P(+|S)} = \frac{0.712}{1 - 0.752} \approx 2.87$$

Así, deberá multiplicar las odds a priori por 2.87 (aproximadamente 3) para obtener las odds a posteriori. Si en cierta consulta privada hay 1 enfermo por cada 3 sanos (*odds a priori*); una vez haya dado positivo el indicador diagnóstico, habrá 1 enfermo por cada 1 sano:

$$odds \text{ a posteriori} = RV \times Odds \text{ a priori} = 2.87 \times 1/3 = 0.96 \approx 1$$

En cambio, si en urgencias hay 7 enfermos por cada sano (*odds a priori* = 7), entre los que den positivo, habrá 20 enfermos por cada sano:

$$odds \text{ a posteriori} = RV \times Odds \text{ a priori} = 2.87 \times 7 = 20.09 \approx 20$$



### Recuerde

*Odds a priori*: razón enfermo/sano propia de un entorno

*Odds a posteriori*: razón enfermo/sano en un entorno tras un resultado positivo

*Razón de verosimilitud*: razón de las probabilidades de positivo entre enfermos y sanos



### Ejercicio 2.5

Suponga que el resultado de la prueba en el ejemplo anterior ha sido negativo, ¿Cuánto vale la RV de un resultado negativo? ¿Cuánto vale la *odds* a posteriori de un resultado negativo?

En resumen, la *odds* conduce a cálculos más simples y directos. Ésta representa su principal ventaja y el motivo de su aplicación en la práctica clínica: En el libro de [Guyat et al.](#) puede ver numerosas aplicaciones clínicas.

**Nota MUY técnica:** Tomando logaritmos, transformamos el producto en una suma:

$$\underbrace{\log(Odds_{posteriori})}_{\text{Información a posteriori}} = \log(LR * Odds_{priori}) = \underbrace{\log(LR)}_{\text{función SOPORTE}} + \underbrace{\log(Odds_{priori})}_{\text{Información a priori}}$$

Es decir, los *logodds* a priori más la información aportada por el resultado empírico (“función soporte”), proporcionan los *logodds* a posteriori.

## 2.4. Tasa

En muchas ocasiones, los casos se observan durante un tiempo variable que conviene tener en cuenta. La tasa incluye en el denominador este tiempo de seguimiento. El riesgo así calculado es el cociente entre un número de eventos y una suma de tiempos de seguimiento, por lo que ya no se trata de una probabilidad (casos posibles entre casos totales).



### Definición

**Tasa** es una relación entre 2 magnitudes. En nuestro caso, la frecuencia de un evento relativa, en general, al tiempo.

**Ejemplo 2.7:** [Regidor E et al.](#) En 1998 se produjeron en España 360511 defunciones, lo que supone una tasa de mortalidad de 915.7 por 100000 habitantes en un año de seguimiento —

o, como generalmente se interpreta, por 100000 habitantes seguidos durante 1 año. En la Tabla 2.2 aparecen ordenadas jerárquicamente, en virtud del número de fallecimientos, las 12 causas de muerte estudiadas.

Causas de muerte (CIEB.*revisión)	Defun- ciones	Tasa	Mortalidad (%)	Porcentaje de cambio	
				1995-98	1980-98
Todas las causas	360511	915.7	100	-3.2	-20.2
Cáncer (140-208)	89665	227.7	24.9	-2.4	7.7
Enfermedades del corazón (390-398. 410-429)	81768	207.7	22.7	-1.8	-26.4
Enfermedad cerebrovascular (430-438)	38121	96.8	10.6	-12.0	-52.8
Enfermedad pulmonar obstructiva crónica (490-496)	17768	45.1	4.9	1.1	30.8
Accidentes no intencionales (E800-E949)	13122	33.3	3.6	0.1	-21.6
Diabetes mellitus (250)	9533	24.2	2.6	-3.4	-15.1
Neumonía e influenza (480-487)	8491	21.6	2.4	3.9	-52.6
Cirrosis y otras enfermedades crónicas del hígado (571)	6246	15.9	1.7	-14.3	-45.3
Nefritis, síndrome nefrótico y nefrosis (580-589)	5566	14.1	1.5	-1.6	-16.1
Aterosclerosis (440)	4717	12.0	1.3	-26.5	-81.2
Enfermedad de Alzheimer (331.0)	3551	9.0	1.0	42.9	2111.6
Suicidio (E950-E959)	3261	8.3	0.9	0.1	53.8

**Tabla 2.2** Principales causas de muerte en España en 1998. Número de defunciones, tasas anuales de mortalidad, mortalidad proporcional y porcentajes de cambio de 1995 a 1998 y de 1980 a 1998. (Modificada)



### Recuerde

El término “por” en la expresión “tantos casos por tantos años de seguimiento” significa “dividido por”, no “multiplicado por”.

Si el riesgo es constante a lo largo del seguimiento, basta con un único valor para representarlo: la tasa de riesgo anterior. Pero si va cambiando a lo largo del seguimiento, necesitamos especificar cuánto vale este riesgo en cada momento del tiempo, lo que llamamos “riesgo en función del tiempo” o, más brevemente, *función de riesgo*. Como analogía, podemos decir que la tasa representa cierta velocidad promedio de aparición de eventos: igual que en un viaje, esta velocidad puede ser más o menos variable. Cuanto menos varíe (“más constante”), más útil será una tasa global; pero cuanto más varíe, más información aporta conocer su valor exacto en cada momento mediante la función de riesgo. En “supervivencia” volveremos a verlos.



### Recuerde

Si el denominador de la medida de frecuencia incluye el tiempo de seguimiento, hablamos de tasa.

## 2.5. Incidencia y prevalencia



### Definición

La **incidencia** estudia el número de casos nuevos durante un período de tiempo.

**Ejemplo 2.8:** [Cohn et al.](#) Sostienen que la incidencia de la combinación de eventos que definían la respuesta de interés fue un 13.2% menor con Valsartán que con placebo.



### Definición

La **prevalencia** estudia el número de casos en un momento de tiempo.

**Ejemplo 2.9:** [Martín et al.](#) La prevalencia de asma varía entre países de Europa, con cifras que oscilan entre un 8% en el Reino Unido y un 2% en Grecia. También existen amplias diferencias entre zonas de un mismo país, encontrándose en España cifras entre el 5 y el 1%.

**Ejemplo 2.10:** [Ricart et al.](#) “El objetivo de este trabajo es determinar la prevalencia de diabetes mellitus gestacional en una población española [...]. Resultados: la prevalencia de DMG fue del 15%”.

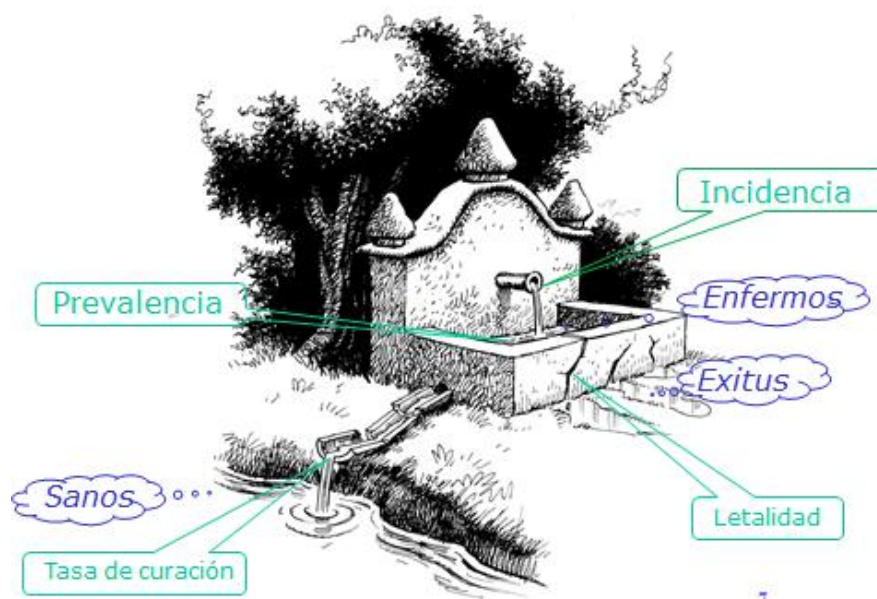


### Recuerde

Incidencia valora casos nuevos y prevalencia, los existentes.

La prevalencia depende de la incidencia, claro, pero también depende de la duración de la enfermedad. Así, aunque la incidencia sea alta, como en la gripe, como la tasa de curación también lo es, la prevalencia será baja. Además de por curación, la duración de la enfermedad puede ser corta por el motivo contrario, la muerte. La tasa de pacientes con la enfermedad que fallece recibe el nombre de letalidad: así, si cierta enfermedad tiene alta letalidad, la prevalencia será baja. Por ello, la incidencia es más informativa en procesos agudos y la prevalencia, en crónicos. Nótese también que la incidencia informa sobre los recursos sanitarios necesarios para afrontar nuevos casos (p.e., quizás urgencias); mientras que la prevalencia, de aquellos necesarios para seguir a los pacientes durante su proceso (p.e., quizás atención primaria). Por otro lado, si se buscan los determinantes de aparición de la enfermedad, la medida de interés es la incidencia.

Una analogía habitual para explicar estos flujos es la fuente de la figura 2.1.



**Figura 2.1:** El flujo que llega a la fuente es la incidencia de nuevos casos, mientras que lo acumulado en el depósito es la prevalencia, que disminuye por las tasas de curación y de letalidad.

### 3. Asociación

Conviene estudiar qué características previas son independientes de la evolución y cuáles están asociadas y pueden, por tanto, ayudar a predecirla.

**Ejemplo 3.1.** Veamos un ejemplo sencillo: el riesgo o probabilidad en los expuestos era de un 5.3% [ $P(Y+|X+) = 7 / 132 \approx 0.053$ ], mientras que en los no expuestos era del 0.9% [ $P(Y+|X-) = 8 / 868 \approx 0.009$ ]: ¿Cuán distintos son 5.3 y 0.9%?

**Nota:** dos números se comparan o mediante su resta o mediante su división.

#### 3.1. Diferencia de riesgos

**Nota:** tradicionalmente se usaba el nombre riesgo atribuible, pero da a entender que este incremento del riesgo viene originado, causalmente, por la exposición al factor; por lo tanto mejor evitarlo para impedir que se interprete que la relación es de causa-efecto.



#### Definición

La **diferencia de riesgos (RA)** es la **diferencia** entre las proporciones de casos en expuestos y en no expuestos



### Recuerde

La diferencia de riesgos debe valorarse únicamente a nivel predictivo: los expuestos tienen ese mayor riesgo que los no expuestos.

**Historieta:** el riesgo de accidente es mayor en los coches rojos. Y aceptamos que los coches rojos paguen un seguro más caro. Pero, para bajar los accidentes, no prohibimos los coches rojos.

**Ejemplo 3.1(continuación):** En el ejemplo anterior, la diferencia entre 0.053 y 0.009 es 0.044; es decir, expresado en porcentajes, un 4.4%.

La diferencia de riesgos puede tomar el valor máximo de 1 (ó 100%) si todos los expuestos desarrollaran la enfermedad y, a la vez, ninguno de los no expuestos.

El valor 0 representa que el riesgo es el mismo en ambos grupos.

**NOTA:** Si la presencia del factor tuviera menos riesgo de enfermedad, se observarían valores negativos (hasta  $-1$  o  $-100\%$ ). Para facilitar la interpretación y expresar los resultados en positivo, basta con intercambiar las definiciones de expuesto y no expuesto.

### 3.2. Número necesario de casos tratados

El objetivo más noble de la Medicina es cambiar el futuro de los pacientes, para lo que recurre a intervenciones clínicas, médicas (como tratamientos farmacológicos, quirúrgicos, fisioterápicos, etc.) o de salud pública, preventivas (como disminuir la exposición a un agente nocivo). Como hemos dicho, estas intervenciones tendrán *efecto*, si (1) la relación es *causal*; y (2) tenemos capacidad de asignación —cambiar una acción por otra, poder *decidir*.

**Ejemplo 3.1(continuación):** si asignamos 1000 casos a X+ y la relación es causal, cabe esperar que 53 casos desarrollen Y+. Si, en cambio, asignamos estos mismos 1000 casos a X-, cabe esperar sólo 9, con un “ahorro” de 44 casos por mil “re-asignados”. Una simple “regla de 3” muestra que si se necesita reasignar 1000 casos para bajar 44 eventos, el Número Necesario de casos a ser Tratados para evitar un evento (NNT o *Number needed to be treated*) sería  $1000/44 = 22.7$ , es decir, unos 23 casos. Recordemos que esta “reasignación” puede referir a un cambio en la exposición (epidemiología); o a un cambio en el tratamiento (clínica).



### Definición:

El número necesario de pacientes a tratar para evitar un evento (NNT) es el inverso del diferencia de riesgos:  $NNT = 1/RA$ .

**Nota:** NNT es muy interpretable, pero que no es tan sencillo si el seguimiento es variable o si cada paciente puede presentar más de 1 evento. Tampoco es cómodo el intervalo de incertidumbre si el de la la diferencia de riesgos pasa por 0.



#### Recuerde

NNT expresa el esfuerzo para evitar 1 evento.

### 3.3 Riesgo relativo



#### Definición

El *riesgo* en los expuestos *relativo* a los no expuestos es el *cociente* entre “riesgo en expuestos” y “riesgo en no expuestos”.

**Ejemplo 3.1(continuación):** la razón entre 0.053 y 0.009 es 6, lo que indica que los expuestos tienen un riesgo 6 veces superior (a los no expuestos).

El riesgo relativo (RR) pretende evaluar cuánto se multiplica la probabilidad de desarrollar la enfermedad. El valor 1 representa que el riesgo es el mismo en ambos grupos. Valores inferiores a 1 indicarían un factor con menor riesgo; y valores superiores a 1, con mayor riesgo.

**NOTA:** Los límites superior e inferior son infinito ( $\infty$ ) y cero (0), para las 2 situaciones extremas en las que no hubiera ningún evento o en los no expuestos, o en los expuestos, respectivamente.



#### Recuerde

El RR valora la razón de riesgos.

**Ejemplo 3.2:** La tabla Tabla 3.1 de [Banegas et al.](#) muestra cómo aumentan los RR de muerte cardiovascular en los pacientes con mayores presiones arteriales respecto al grupo de referencia formado por los individuos con menores valores, “<120/80” —al que, por definición, le corresponde un  $RR = 1$ . La interpretación del grupo con mayores presiones es: un 3.3% de hombres tuvo cifras >180/100 y presentaron el evento con una frecuencia 3.4 veces superior que el grupo de referencia (tuvieron un 340% de eventos más que el grupo de menor presión, <120/80). Es decir, el grupo de presión más alta tiene una probabilidad más de 3 veces mayor de presentar un evento que el grupo de menor presión.

Presión arterial (mmHg)	Hombres	
	%	Todas las causas (RR)
<120/80	20.1	1
120-129/80-84	18.1	1.2
130-139/85-89	17.2	1.3
140-155/90-99	29.1	1.6
160-169/100-109	17.2	2.2
>180/110	3.3	3.4
TOTAL	100	

**Tabla 3.1** Prevalencias de presión arterial en España y RR de muerte cardiovascular.

### 3.4 Odds ratio o razón de momios



#### Definición

El *odds ratio* es el cociente entre la *odds* en los expuestos y la misma *odds* en los no expuestos.

**Ejemplo 3.1(continuación):** la razón entre 0.056 y 0.009 es 6.2, indicando que la *odds* en los expuestos es 6.2 veces superior a la *odds* en los no expuestos.

Como hemos dicho, si la enfermedad es poco frecuente, *odds* y probabilidad son similares y por tanto también *odds ratio* y riesgo relativo, en este ejemplo 6.2 y 6. Ambas medidas se interpretan de forma similar.



#### Ejercicio 3.1

Calcule RR, RA y OR en las tablas del **Ejercicio 2.4**. A pesar de que se ha dicho que no tenía sentido calcular los riesgos ni las *odds* en la tercera tabla, haga también en ella todos estos cálculos y observe qué sucede con el valor del *odds ratio*.

Una gran ventaja del *odds ratio* sobre las medidas basadas en riesgos es que puede ser calculado en cualquier tabla 2x2, independientemente del plan de muestreo.

**Nota técnica:** ello es así porque, de la misma forma que se definió el OR como el cociente entre expuestos y no expuestos de las *odds* enfermo / sano  $[(a/b)/(c/d)]$  también podría haberse definido como el cociente entre enfermos y sanos de las *odds* expuesto / no expuesto  $[(a/c)/(b/d)]$ . Dado que ambas definiciones son equivalentes, el *odds ratio* se puede utilizar en cualquier tabla de dos filas y dos columnas, independientemente del plan de muestreo:  $(a/b)/(c/d) = (a/c)/(b/d) = ad/bc$  [*a*, *b*, *c* y *d* representan los cuatro valores de una tabla 2x2 – con exposición en filas y enfermedad en las columnas – de arriba abajo y de izquierda a derecha] .



**Recuerde:**

El *odds ratio* tiene la gran ventaja de que se puede utilizar en cualquier estudio, independientemente del plan de muestreo.

**Ejercicio 3.2**

Imagine un diseño casos-contróles, en los que se escoge una muestra de enfermos (casos) y una muestra de sanos (contróles), por lo que se deja fija la variable enfermo / sano. ¿Cuáles de las medidas anteriores (RR, RA o OR) pueden aplicarse?

**Ejercicio 3.3**

¿Qué relación existe entre el riesgo relativo y el odds ratio?

**Recuerde:**

No puede calcular riesgos, ni su diferencia, ni su cociente en estudios casos-control.

**3.5 Razón de tasas (*Hazard ratio* o HR)**

Para comparar las tasas de eventos, recurrimos a su cociente, al que llamamos razón de tasas o *hazard ratio* (HR). Por ejemplo,  $HR=0.8$  significa que la probabilidad instantánea de morir en el grupo tratado equivale al 80% de dicha probabilidad en el control: asignar al tratamiento evita un 20% de muertes en cada instante.

Nótese que la muerte no se puede evitar, tan sólo atrasar, lo que hace muy delicada la interpretación del HR. En los puntos siguientes daremos 2 pistas para su interpretación, primero comentado que los valores del HR ocupan un lugar intermedio entre los de OR y RR; y segundo, facilitando su relación con el incremento de vida.

Antes, recopilemos las medidas de riesgos. La Tabla 3.2 resume las medidas más usuales, distinguiendo por columnas entre las que se emplean para describir 1 grupo y las que se usan para comparar 2 grupos. Por filas, distingue entre estudios transversales y longitudinales; y estos últimos, según si el tiempo de seguimiento es fijo o variable.

		Descripción 1 grupo		Comparación 2 grupos	
Prevalencia (casos existentes)	Puntual (un instante)	Riesgos	Odds	Riesgo relativo (RR) Diferencia de riesgos (RA) Número necesario (NNT)	Odds ratio (OR)
Incidencia o mortalidad (nuevos casos)	Seguimiento idéntico				
	Seguimiento variable	Tasa de riesgo o Hazard Rate (constante)  Función de riesgo o Hazard Function (variable)	Razón de riesgos o Hazard Ratio (HR)		

**Tabla 3.2.** Medidas epidemiológicas más relevantes.



### Ejercicio 3.4

Al decidir usar una tasa en lugar de un riesgo o una odd, la clave es...

Si el riesgo es constante a lo largo del seguimiento, basta con un único valor para representarlo: es la *tasa de riesgo* que ya hemos visto antes. En cambio, si va cambiando a lo largo del seguimiento, necesitamos especificar cuánto vale este riesgo para los diferentes momentos del tiempo: lo llamamos *función de riesgo*.

**Analogía:** La tasa representa la velocidad de aparición de eventos: igual que un viaje, esta velocidad puede ser variable o constante. Una tasa global puede ser útil, pero si la tasa instantánea varía, conocerla aporta más información.



### Ejercicio 3.5

La clave para usar una tasa o una función de riesgo es ...

**Nota:** El punto de supervivencia (3) del capítulo 11 estudia más a fondo el HR.

Desgraciadamente, los artículos suelen ser creativos en el término empleado para referirse a estas 3 medidas. Si el estudio que Vd. desea interpretar usa un término parecido pero diferente, recurra al artículo de [Lisa Schwartz](#) para ver su equivalencia con las medidas comentadas.

## 3.6 Relación entre los 3 cocientes de riesgos: RR, HR y OR \*

**NOTA:** Algunos autores emplean el término riesgo relativo para estas 3 medidas: RR, HR y OR.

Existe una gradación entre sus valores: OR muestra siempre mayores valores de relación que HR; y éste mayores que RR.

**Nota técnica:** [Symons MJ et al.](#) Como valores ‘mayores’ debe interpretarse como más alejados de 1, sea por encima o sea por debajo, la expresión formal de esta desigualdad es:

$OR \leq HRR \leq RR \leq 1$  para factores con menor riesgo (p.e., intervenciones)

$1 \leq RR \leq HRR \leq OR$  para factores con mayor riesgo (p.e., exposiciones)



### Ejercicio 3.6

¿Qué medida ‘realza’ el efecto de una intervención?

Los tres proporcionan valores más similares cuanto más pequeños son (1) el riesgo considerado, (2) el efecto diferencial y (3) el periodo de seguimiento.

Dada esta similitud, algunos autores emplean de forma amplia los términos riesgo relativo y *risk ratio* para referirse tanto al RR, como al OR o el HR.



### Recuerde:

El cociente de riesgos más realza la relación es OR y la que menos, RR, quedando HR entre ellas.

## 3.7 Relación entre HR y las medianas de los tiempos hasta el evento.

El HR, en ciertas condiciones teóricas especiales, se corresponde con el cociente de las medianas de los tiempos de vida. Así, tomando la mediana como un estimador de la esperanza de vida, en el ejemplo anterior de  $HR=0.8$ , diríamos que la esperanza de vida en los pacientes del grupo control es un 80% de la de los tratados, es decir, un 20% inferior. Desgraciadamente, esta interpretación no puede aplicarse siempre, aunque puede ser una primera [aproximación](#) para valorar el posible beneficio de una intervención.



### Recuerde:

En ciertas ocasiones HR equivale a cocientes de esperanzas de vida

## Soluciones a los ejercicios

**1.1** Dado que  $P(B|A)$  debe representar la probabilidad del suceso B ‘dentro’ del total de casos que cumplen A, se trata de dividir la probabilidad de ser a la vez A y B, es decir  $P(A \cap B)$ , por la probabilidad de ser A, es decir  $P(A)$ . En resumen, tiene el mismo numerador que  $P(A|B)$ , pero cambia el denominador. 
$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Gráficamente sería el cociente entre la zona  $A \cap B$  y la zona A.

**1.2** Es más grande  $P(M|C)$ , dado que  $P(M|C)$  representar la probabilidad de ser Mujer del total de casos que tienen cáncer de Mama.

**1.3**  $0.25 = P(\text{Sí}|M) = P(\text{Sí}|H)$

**1.4**  $\text{Sens} = P(+|E)$  Proporción de positivos en el conjunto de enfermos.

$\text{Esp} = P(-|S)$  Proporción de negativos en el conjunto de sanos.

$\text{VP+} = P(E|+)$  Proporción de enfermos en el conjunto de positivos.

$\text{VP-} = P(S|-)$  Proporción de sanos en el conjunto de negativos.

**1.5** La tabla correspondiente es:

	+	-	Total
Enfermo	132		132
Sano		868	868
Total	132	868	1000

$$\text{VP+} = P(E|+) = 132 / 132 = 1.00 = 100\%$$

$$\text{VP-} = P(S|-) = 868 / 868 = 1.00 = 100\%$$

**1.6** Para que  $\text{VP+} = 100\%$  se requiere que los casos positivos provengan todos de los enfermos, es decir: que ningún sano dé positivo. Por tanto, se requiere que la especificidad sea del 100% para tener un  $\text{VP+}$  del 100%. [Similarmente, para que el  $\text{VP-}$  sea del 100% se requiere que la sensibilidad sea del 100%.]

**1.7** Una prueba sería muy-muy específica si, aplicada a un conjunto de sanos, casi el 100% dan negativo y muy-muy sensible si, aplicada a un conjunto de enfermos, casi en su totalidad dan positivo. Un resultado tendrá un elevado valor predictivo negativo si casi el 100% de los que dan negativo están realmente sanos y un valor predictivo de casi el 100% si de los que dan positivo, casi todos están enfermos.

**1.8** La especificidad y el valor predictivo positivo ( $\text{VP+}$ )

**1.9**  $\text{Sens} = P(+|E) = 712/1000 = 0.712 = 71.2\%$

$$\text{Esp} = P(-|S) = 752/1000 = 0.752 = 75.2\%$$

$$\text{VP+} = P(E|+) = 712/960 = 0.741 = 74.1\%$$

$$\text{VP-} = P(S|-) = 752/1040 = 0.723 = 72.3\%$$

$$1.10 \quad VP- = P S - = \frac{P - S \cdot P S}{P - S \cdot P S + P - E \cdot P E} = \frac{0.752 \cdot 0.868}{0.752 \cdot 0.868 + 0.288 \cdot 0.132} = 0.945$$

### 1.11

Población 1

$$VP+ = P E + = \frac{P + E \cdot P E}{P + E \cdot P E + P + S \cdot P S} = \frac{0.95 \cdot 0.8}{0.95 \cdot 0.8 + 0.05 \cdot 0.2} = 0.9870$$

Población 2

$$VP+ = P E + = \frac{P + E \cdot P E}{P + E \cdot P E + P + S \cdot P S} = \frac{0.95 \cdot 0.01}{0.95 \cdot 0.01 + 0.05 \cdot 0.99} = 0.1610$$

2.1 La tabla correspondiente es:

	AA: Y+	No AA: Y-	Total
EFV:	36	15	51
PI:	9	40	49
Total	45	55	100

Riesgo:  $P(Y+) = 45/100 = 0.45 = 45\%$

Riesgo en los Expuestos a EFV:  $P(Y+|EFV) = 36 / 51 \approx 0.71 = 71\%$

Riesgo en los Expuestos a PI:  $P(Y+|PI) = 9 / 49 \approx 0.18 = 18\%$

2.2 La probabilidad de sacar un “3” vale  $1/6=0.167=16.7\%$ . La *odds* respectiva es  $1/5=0.2=20\%$ . En una casa de apuestas se pagaría 5 euros por euro apostado.

2.3 *Odds* (dolor lumbar) =  $P(\text{dolor lumbar}) / P(\text{no dolor lumbar}) = 0.5 / 0.5 = 1$  (la *odds* de padecer dolor lumbar a lo largo de la vida están “1 a 1”).

*Odds* (gripe) =  $0.8 / 0.2 = 4$  (la *odds* de gripe están 4 a 1: por cada persona que no padecerá gripe, hay 4 que si la tendrán).

*Odds* (esclerosis múltiple) =  $0.001 / 0.999 = 0.001001001 \approx 0.001$  (la *odds* de esclerosis múltiple está 1 a 1000).

2.4 Los resultados que figuran a continuación muestran que los riesgos y las *odds* de la última tabla no coinciden con los anteriores. ¿Qué ha pasado? Nótese que la variable respuesta (bronquitis) de la tercera tabla ya no depende de las observaciones, pues, por diseño, se ha dejado fija: se ha construido una tabla que tiene, porque así lo hemos querido, la mitad de bronquíticos y la mitad de no bronquíticos. Estos datos, obtenidos de dos muestras de la variable respuesta ya no sirven para calcular ni los riesgos ni las *odds* de desarrollar una bronquitis. Nótese que en la segunda tabla, en la que había una muestra de fumadores y una de no fumadores sí que podían calcularse los riesgos y las *odds* condicionados a fumador o a no fumador.

	Riesgos			Odds		
	Tabla 1	Tabla 2	Tabla 3	Tabla 1	Tabla 2	Tabla 3
En los expuestos: X+	0.71	0.71	0.85	2.47	2.47	5.53
En los no expuestos: X-	0.25	0.25	0.42	0.33	0.33	0.74

2.5 Para calcular el RV de un resultado negativo procederemos:

$$RV = \frac{P(-|E)}{P(-|S)} = \frac{1-0.712}{0.752} = 0.38$$

Con la misma proporción de enfermos tenemos que:

$$Odds \text{ a posteriori} = RV \cdot Odds \text{ a priori} \rightarrow Odds \text{ a posteriori} = 0.38 \cdot 1/3 = 0.13$$

3.1 Como cabía esperar, la tercera tabla, en la que se había dejado fijo el número de casos con y sin la enfermedad, no permite calcular ni el riesgo relativo ( $2.87 \neq 2.00$ ) ni la diferencia de riesgos ( $0.46 \neq 0.42$ ). Pero sí que permite calcular el *odds ratio*: se obtienen los mismos valores (7.51) ya que el *odd ratio* puede obtenerse con cualquier diseño, independientemente de las restricciones impuestas en el diseño.

	Riesgos			Odds		
	Tabla 1	Tabla 2	Tabla 3	Tabla 1	Tabla 2	Tabla 3
En los expuestos: X+	0.71	0.71	0.85	2.47	2.47	5.53
En los no expuestos: X-	0.25	0.25	0.42	0.33	0.33	0.74
Diferencia de riesgos	0.46	0.46	0.42			
Riesgo relativo	2.87	2.87	2.00			
<i>Odds ratio</i>				7.51	7.51	7.51

3.2 Por lo dicho anteriormente, sólo el *odds ratio* (OR). El riesgo relativo y el diferencia de riesgos no tienen valor en este tipo de diseños ya que sus valores dependerán del número de controles que se haya decidido seleccionar para cada caso.

3.3 Son estimadores de la relación entre dos variables dicotómicas, una supuesta respuesta y una supuesta causa. Cuando la proporción de la respuesta es muy pequeña, dan valores muy parecidos. Se interpretan igual.

3.4 Que el seguimiento de los pacientes sea variable.

3.5 Una única tasa global sólo escogerá toda la información cuando el riesgo instantáneo se mantenga constante durante el periodo considerado.

3.6 El *Odds ratio* proporciona valores mayores, por lo que deberá estar atento: no es que hayan trampa ni sesgo, pero es una medida diferente que proporciona valores más extremos. Y este efecto es más exagerado cuanto mayor sea la frecuencia del evento considerado.